An Automated Framework for Railway Sleeper Detection and Spacing Measurement using UAV Imagery

Bhanu Pratap Singh[1] and Sahil Garg[2]

[1]Department Of Civil and Environmental Engineering, Indian Institute of Technology Delhi, New Delhi, India
[2]Department Of Civil and Environmental Engineering, Indian Institute of Technology Delhi, New Delhi, India
cec242814@iitd.ac.in, gargsahil@iitd.ac.in

Abstract –

Railway sleepers form the structural interface between the rails and the ballast, governing track geometry stability through load transfer and positional restraint, making them a critical track component. Regular monitoring of sleeper condition and spacing contributes to maintaining track condition and supports safe railway operation for goods and passenger traffic. In current practice, sleeper spacing inspection on ballasted track relies on manual tape measurements and visual judgement, which are time consuming, labour-intensive, and difficult to apply consistently over long construction sections or during short night possessions.

This paper proposes an automated approach for sleeper spacing measurement using unmanned aerial vehicle imagery and a hybrid computer vision-based processing framework. The method applies a YOLOv11m object detection model to localise individual sleepers in near-nadir UAV images, followed by mask refinement using the Segment Anything Model to delineate sleeper regions. Sleeper centres and rail edges are extracted from the resulting masks, and pixel distances are converted to physical spacing through calibration based on the known effective track gauge.

The framework was evaluated on a dataset of straight ballasted track sections acquired under routine field conditions. Quantitative validation demonstrated a mean absolute error of 6.56 mm and a root mean square error of 7.67 mm, with 93.75% of predicted spacing values falling within ±10 mm of reference measurements. These results indicate that the proposed workflow enables reliable and physically interpretable sleeper spacing measurements from UAV imagery.

Keywords –
Ballasted Track, UAV-Based Inspection, YOLO-Based Object Detection, Image Segmentation, Sleeper Spacing.

1 Introduction

Railway networks rely on consistent track geometry to maintain safe operations, control dynamic wheel-rail interactions, and limit long-term degradation of sleepers, ballast, and fastening systems. Within this broader geometry envelope, the longitudinal spacing of sleepers plays a very practical role in how loads are distributed, how ballast consolidates, and how maintenance limits are defined in track standards and inspection manuals [1]. In day-to-day practice, however, sleeper spacing for ballasted track is still often checked through manual gauge and visual inspections or only at discrete locations using track recording cars, which can miss localised anomalies, construction non-conformities, or gradual shifts that emerge between scheduled runs [2]. For railways that are expanding or upgrading lines with limited track access, this combination of manual point-based measurements and visual checks is increasingly difficult to reconcile with the demand for more frequent, fine-grained quality control.

Image-based inspection offers a potential route toward denser geometric assessment. An ideal scenario would use nadir images of the track acquired from a UAV or similar platform under controlled conditions, with known camera parameters and stable flight geometry. In such a scenario, each sleeper could be detected, cleanly segmented from ballast and rails, and used together with calibrated reference dimensions to compute spacing in physical units in a transparent and interpretable manner. In practice, the conditions are far less controlled. UAV imagery is affected by variable altitude, camera tilt, motion blur, and changing lighting; tracks themselves may present occlusions due to vegetation, ballast pockets, or construction materials; and, in many projects, full camera calibration or ground control is either unavailable or only loosely specified. These factors make it difficult to move from visually plausible sleeper detection to reliable metric spacing measurements that can be trusted for acceptance and compliance checks [3], [4].

Recent vision-based railway work has advanced along several strands that do not always converge. Object detection models, particularly YOLO variants, identify sleepers, fasteners, and surface defects at high frame rates in inspection vehicles or UAVs, prioritising throughput, and automation rather than geometric boundary precision [5], [6]. UAV imagery has supported track localisation, ballast assessment, and overhead line inspection, showing aerial data can integrate into workflows when operational constraints are managed. Visual metrology systems for gauge and other parameters achieve sub-millimetre to millimetre-scale engineering accuracy under careful calibration, but typically with fixed cameras, dedicated targets, and constrained environments rather than mobile UAV platforms.

Sleeper-level geometry highlights the trade-off between flexibility and rigour. UAV-based approaches using template matching or simple image processing report centimetre level spacing estimates, but over limited, visually clean segments and under stable scale/pose assumptions. Higher accuracy gauge or profile systems rely on controlled illumination, rigid setups, or track-mounted artefacts that are difficult to reproduce over long aerial surveys. This leaves a gap between fast, detection-centric pipelines treating sleepers as coarse pixel objects and metrology-centric systems handling geometry carefully but in constrained configurations. Despite these developments, a clear understanding of how detection-driven segmentation workflows perform as measurement-oriented tools under practical UAV inspection conditions remains limited.

The present study addresses this gap by focusing explicitly on the geometric reliability of sleeper spacing measurement from UAV imagery using a hybrid detection and segmentation framework. Rather than proposing a new detection architecture, the contribution lies in evaluating whether segmentation-refined detections combined with explicit geometric calibration can produce spacing outputs that are geometrically consistent and practically interpretable under typical UAV inspection conditions. The framework employs a YOLO-based model to localise sleepers in near-nadir UAV images, followed by the Segment Anything Model to refine detections into detailed masks. This refinement improves boundary definition and supports reliable centre estimation for spacing measurement. In addition, the workflow integrates an explicit calibration strategy using track gauge as a geometric reference, enabling conversion from pixel distances to physical units without requiring external ground control. The study focuses on defined straight track sections to establish a controlled baseline and assess whether such an integrated approach can support spacing measurement for construction verification and routine maintenance inspection. This study evaluates the reliability of a hybrid detection and segmentation workflow for sleeper spacing measurement.

2 Related Work

Vision-based inspection has become an important approach in railway infrastructure monitoring, enabling large networks to be inspected with reduced reliance on manual surveys and specialised geometry vehicles. Camera-based systems mounted on inspection vehicles, wayside installations, and mobile platforms are widely used to detect rails, sleepers, fasteners, and ballast conditions. Many of these systems rely on deep learning-based object detection models, particularly YOLO variants, which prioritise rapid localisation and automation rather than precise geometric boundary representation [5], [6].

UAV imagery has further expanded inspection coverage and deployment flexibility, particularly in construction zones and remote locations. Prior studies demonstrate the use of UAV data for track localisation, ballast assessment, and geometric parameter estimation such as track alignment or gauge using photogrammetric or model-based approaches [7]. While these methods confirm the feasibility of metric measurement from aerial imagery, they often rely on dense ground control, careful mission planning, or computationally intensive three-dimensional reconstruction, which limits their practicality for routine sleeper spacing measurement over short track sections.

A parallel body of work focuses on image-based rail geometry and metrology. Visual measurement systems for track gauge and wheel rail position show that millimetre level accuracy is achievable when camera calibration, reference geometry, and error modelling are treated explicitly [8]. However, such systems typically operate in controlled environments with fixed or stereo camera setups and do not readily transfer to UAV imagery acquired under variable pose and lighting conditions. Deep learning models such as YOLO and Mask R-CNN-based segmentation has been applied to estimate inter-sleeper distances from image data, demonstrating feasibility under controlled conditions [9].

Taken together, existing studies either prioritise detection performance without explicit treatment of geometric interpretation or achieve precise measurements under controlled configurations that are difficult to reproduce in routine UAV operations. As a result, the geometric reliability of sleeper spacing measurement derived from overhead imagery, particularly when calibration resources are limited and track gauge is the primary physical reference, remains insufficiently understood. This limitation motivates the present study, which evaluates whether segmentation-refined detection combined with practical geometric calibration can support reliable sleeper spacing measurement under realistic field conditions.

3 Data Acquisition and Study Area

UAV imagery used in this study was collected along straight ballasted railway sections in the Mau-Aunrihar corridor of the Varanasi Division, India, using a DJI Mavic 3 unmanned aerial vehicle equipped with an onboard RGB camera. Image acquisition was carried out in a near-nadir configuration to minimise perspective distortion and maintain a consistent geometric view of sleeper placement along the track. Straight sections were selected intentionally to isolate geometric reliability before extension to curved alignment.

All images were captured from a near-nadir viewpoint under natural daylight conditions without artificial markers, calibration targets, or ground control points. Natural variations in ballast distribution, sleeper surface condition, and surrounding track environment were retained to reflect routine field inspection conditions. Images were stored at full resolution and processed without geometric rectification beyond standard onboard stabilisation. Track gauge, clearly visible in the imagery, was used as the sole physical reference for pixel-to-millimetre scale conversion.

Table 1. Dataset Characteristics

Parameter	Description
UAV Platform	DJI Mavic 3
Imaging Orientation	Near-Nadir (Top-Down)
Flight Altitude	12-15 m
Image Resolution	4000 × 2250 pixels
Track Geometry	Straight, Ballasted Track
Track Coverage	Two parallel tracks
Sleeper Type	Concrete Sleepers
Scale Reference	Track Gauge

For this study, two hundred images were selected for experimental evaluation. The imagery was collected along an approximately 4 km operational railway stretch including both open track and yard regions. These environments introduced variation in ballast condition, sleeper surface wear, shadow intensity, and minor vegetation presence. Two different concrete sleeper types were present within the surveyed section, introducing additional geometric variability.

While the dataset was obtained from a single operational corridor, the extended spatial coverage, environmental diversity, and presence of multiple sleeper types provide a representative baseline for evaluating geometric measurement reliability under realistic field variability. Further validation across multiple corridors and track geometries will be required to establish full operational generalisation.

3.1 Data Pre-Processing and Annotation

Annotation was performed manually using rectangular bounding boxes for two object classes, sleeper and rail. Bounding box annotation was selected to ensure reliable localisation of sleeper regions rather than precise boundary delineation. Partially obscured sleepers affected by ballast, shadows, or surface wear were annotated using visible portions. Rail annotations were included only to support geometric reference and track gauge estimation in later processing stages.

The annotated dataset consisted of 200 images divided into training, validation, and test subsets using a 70 percent, 20 percent, and 10 percent split, respectively. This partitioning enabled model training, parameter tuning, and independent performance evaluation. Figure 1 presents representative annotated UAV images.

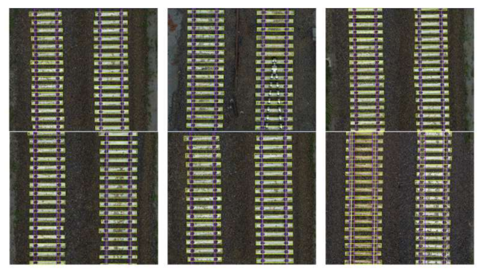

Figure 1. Sample of Annotated UAV Images.

To improve robustness under operational variability, the training data were augmented using controlled geometric, photometric, and noise-based transformations, as summarised in Table 2. After augmentation, the effective size of the training dataset [12] increased to 496 images.

Table 2. Augmentation applied to the training dataset.

Parameter	Range	Purpose
Rotation	−15° to +15°	To handle slight changes in camera orientation
Hue	−10° to +10°	To account for colour changes caused by lighting
Brightness	−15% to +15%	To capture differences in illumination
Gaussian blur	up to 0.5 px	To represent slight blur in some images
Random noise	≤ 0.1% pixels	To reflect minor noise during image capture

4 Methodology

This section describes the methodology for automated railway sleeper detection and spacing measurement using UAV imagery. The proposed hybrid framework integrates detection, segmentation, and geometric processing to derive sleeper spacing from single-view nadir images using track gauge as the physical reference. The workflow is designed to ensure reliable localisation under variable visual conditions and consistent geometric handling prior to spacing estimation.

The framework is designed with two central considerations. First, reliable sleeper localisation must be achieved under variable visual conditions, including ballast overlap, shadows, surface wear, and changes in illumination. Second, spacing estimation requires explicit geometric handling of sleeper positions rather than direct use of object detection outputs. To address these requirements, detection and segmentation are integrated with a structured geometric processing pipeline that supports consistent centre estimation and longitudinal alignment prior to spacing measurement.

4.1 Overview

The proposed framework consists of four important sequential stages: sleeper detection, segmentation and mask refinement, sleeper centre estimation and alignment, and sleeper spacing measurement. The UAV image serves as the input, and spacing values are obtained independently for each track present in the scene. Track gauge is used as the sole physical reference for image (pixel)-to-ground scale conversion.

Deep learning models such as YOLO and the Segment Anything Model (SAM) are widely used for object detection and instance segmentation in complex visual environments. These models learn hierarchical [13] feature representations directly from image data, enabling robust localisation and boundary extraction under the variability present in UAV-based railway imagery.

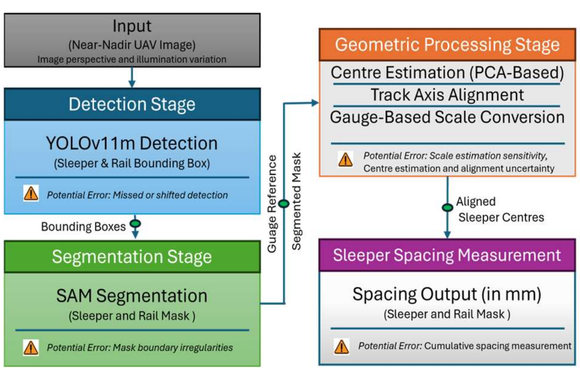

Figure 2. Proposed workflow illustrating detection, segmentation, geometric processing, and spacing measurement stages, with representative sources of uncertainty contributing to cumulative measurement error.

The framework adopts a hybrid deep learning approach that combines a single stage object detector with an instance segmentation model. YOLO operates on a convolutional neural network based single stage detection architecture to provide efficient coarse localisation of railway sleepers. SAM operates on a transformer-based segmentation architecture and uses detector outputs as prompts to generate refined instance level masks. An overview of the complete processing workflow is illustrated in Figure 2.

Each stage of the workflow introduces positional uncertainties that propagate into the final spacing measurement. Detection errors may slightly shift bounding box placement, while segmentation inaccuracies caused by ballast overlap or shadows influence boundary definition. These effects impact centre estimation and alignment, and any error in rail localisation affects the pixel-to-metric scaling process. The influence of these uncertainties is reflected in the spacing error statistics reported in Section 5.3.

4.2 Detection Model Training

Nvidia Tesla T4 GPUs with 16 GB of memory were utilised for model training within the Kaggle notebook environment. The YOLOv11m architecture was employed, initialised with pretrained weights from the COCO dataset. Using the annotated dataset described in Section 3.1, training was conducted on high resolution UAV imagery originally captured at 4000×2250 pixels. During training, images were resized to 1024×1024 pixels for computational consistency. Bounding box annotations were provided for two object classes: sleeper and rail.

The model was trained for 100 epochs with a batch size of 4 using the AdamW optimiser and an initial learning rate of 1×10^{-4}. Early stopping was enabled based on validation loss stabilisation to prevent overfitting. Data augmentation included controlled geometric and photometric transformations. The confidence threshold during inference was set to 0.25, and non-maximum suppression was applied with an intersection over union threshold of 0.50. Model performance was monitored using validation loss and mean average precision during training to ensure stable convergence.

Within the overall processing pipeline, the role of the detection model was intentionally limited. Detection outputs were not used directly for sleeper spacing measurement or geometric analysis. Instead, sleeper bounding boxes were used to initialise subsequent segmentation, while rail detections provided geometric references for track gauge estimation and track axis definition. All geometric accuracy related to sleeper centre estimation, alignment, and spacing measurement was handled in later processing stages.

4.2.1 SAM Vit-H Configuration

The Segment Anything Model (SAM Vit-H) was used in inference mode with pretrained weights and without additional fine-tuning. Bounding box detections generated by the YOLOv11m model were used as prompt inputs to initialise segmentation for each detected sleeper instance. Standard inference configuration was adopted to ensure reproducibility, with automatic mask generation applied to each prompted region. Predicted masks were filtered by retaining the largest connected component to remove small-isolated regions and noise. This post-processing step ensured consistent boundary representation for subsequent centre estimation and spacing measurement.

4.3 Sleeper Detection

In the first stage, individual sleepers are localised using a YOLOv11m-based object detection model trained on domain specific UAV imagery of railway tracks. Bounding box annotations for sleepers were prepared to support detector preparation. The detector processes the complete image and outputs bounding boxes corresponding to sleeper instances across both tracks. No assumptions are made regarding uniform spacing, sleeper orientation, or symmetry between tracks. The bounding boxes generated at this stage provide coarse spatial localisation, and the identified image regions are then analysed in the next stage. They are not used directly for geometric measurement; instead, segmentation is performed using the regions identified in the detection stage.

4.3.1 Rail Localisation for Geometric Reference

Rail instances were detected to establish track gauge and define the longitudinal track axis required for centre alignment and spacing computation. Rail outputs were used only for geometric reference.

4.4 Segmentation and Mask Generation

Segmentation was applied to refine the initial sleeper detections using the Segment Anything Model (SAM Vit-H). Detector outputs were used as prompts to generate binary masks representing sleeper regions. Segmentation improves boundary definition compared to bounding boxes and provides geometrically consistent regions for centre estimation. The resulting masks were used directly in subsequent geometric processing.

4.5 Sleeper centre estimation and alignment

For each segmented sleeper, a centre point is derived from the spatial distribution of its mask pixels, after which principal component analysis (PCA) is used to estimate the sleeper's longitudinal axis. The sleeper centre is defined as the midpoint along this axis.

To ensure consistent longitudinal measurement, the centres of individual sleepers are projected onto an estimated track axis derived from their overall arrangement along each rail. This alignment step reduces lateral variation caused by perspective effects, minor detection inaccuracies, and sleeper skew, and simplifies the spacing analysis by representing sleeper positions along a single one-dimensional track direction.

Following alignment, each sleeper centre is represented by a single scalar coordinate s_i corresponding to its projected position along the estimated track axis. Sleepers are then ordered sequentially along this axis, with indexing performed separately for the left and right tracks.

4.6 Sleeper spacing measurement

Following centre estimation and alignment, sleeper spacing is computed independently for each track using the one-dimensional representation established in Section 4.5. After alignment, each sleeper centre is characterised by a scalar coordinate s_i, representing its projected position along the estimated track axis in pixel units. Sleepers are ordered sequentially along this axis for the left and right tracks separately.

The sleeper spacing is defined as the longitudinal separation between the aligned centres of two consecutive sleepers. Accordingly, the pixel domain spacing between sleeper i and sleeper i+1 is computed as

$$d_i^{px} = s_{i+1} - s_i \quad (1)$$

To express the spacing in physical units, a pixel-to-metric scale factor is derived using the known nominal track gauge visible in the image. Let the physical track gauge be denoted by G_{mm} (in millimetres) and let G_{px} denote the corresponding gauge measured in pixel units from the detected rail positions. The pixel-to-metric conversion factor is then defined as

$$\alpha = \frac{G_{mm}}{G_{px}} \quad (2)$$

The sleeper spacing in real-world units is obtained by applying this scale factor to the pixel-domain spacing as

$$d_i^{mm} = \alpha d_i^{px} \quad (3)$$

Track gauge is used as the physical reference for scale conversion, enabling spacing values to be expressed in real world units without reliance on external ground control points, calibration targets, or multi-view reconstruction.

The complete sequence of processing stages, from (a) sleeper detection, (b) segmentation, (c) centre alignment, and (d) spacing measurement, is shown in Figure 3.

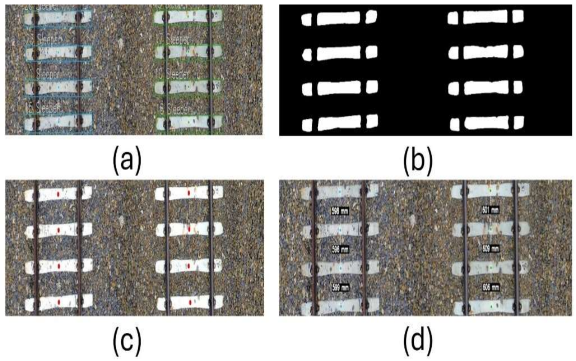

Figure 3. Stages of the proposed framework

5 Result and Discussion

5.1 Detection Performance Evaluation

The detection component of the framework was evaluated to verify that sleepers and rails could be identified reliably for use in later segmentation and spacing analysis. Performance was assessed on the test dataset using precision, recall, and mean average precision at an intersection over union threshold of 0.50.

Intersection over union is used to quantify the spatial agreement between a predicted bounding box B_p and its corresponding ground truth annotation B_g by comparing the area of overlap to the total area covered by both regions. It is defined as

$$IoU = \frac{area\ B_p \cap B_g}{area\ B_p\ \cup B_g} \tag{4}$$

Based on a predefined IoU threshold, detections are categorised as correct or incorrect, forming the basis for computing precision and recall following standard object detection evaluation practice [10]. Precision reflects the proportion of predicted objects that correspond to valid detections, while recall represents the proportion of ground truth objects that are successfully identified by the model [13] and are defined as

$$p = \frac{TP}{TP + FP} \tag{5}$$

$$r = \frac{TP}{TP + FN} \tag{6}$$

where TP, FP and FN denote true positive, false positive, and false negative detections, respectively. Mean average precision provides an overall measure of detection reliability across confidence thresholds and object classes [10], [11].

Performance on the test dataset shows a precision of 93.4% and a recall of 90% for the two object classes. The corresponding mean average precision at an intersection over union threshold of 0.50 is 95% [14]. With a recall of 0.90, the model captured nearly all sleeper and rail instances visible in the test imagery. These results confirm that the detection stage provides consistent and repeatable object localisation under near-nadir UAV imaging conditions.

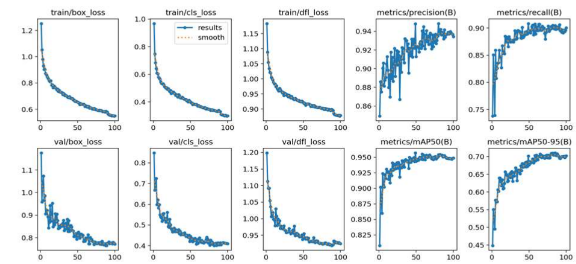

Figure 4. Training performance of the YOLOv11m detection model showing loss convergence and evolution of mean average precision across training epochs.

Figure 4 illustrates the training behaviour of the detection model, showing convergence of training and validation losses together with the progression of mean average precision across epochs.

Sleeper detection outputs are shown in Figure 3(a), where bounding boxes are overlaid on sleepers across both tracks in a near-nadir UAV image. The detections demonstrate consistent localisation across the image extent and provide reliable initialisation for subsequent segmentation and geometric processing stages.

5.2 Comparative Evaluation of Segmentation Models

To provide comparative context for segmentation performance, the proposed YOLOv11m-SAMvit-H workflow was evaluated alongside a Mask R-CNN instance segmentation model trained using the same dataset partitions. The Mask R-CNN model was trained for 50 epochs using the same GPU hardware and notebook environment as the YOLO-SAM implementation to ensure consistent training conditions.

The YOLO-SAM workflow achieved a mean average precision of 70% across IoU thresholds from 0.50 to 0.95, with a mean average precision of 92.78% at IoU = 0.50. The corresponding precision and recall values were 91% and 86%, respectively.

The Mask R-CNN model achieved a mean average precision of 64.2% across IoU thresholds from 0.50 to 0.95, with a mean average precision of 95% at IoU = 0.50 and a recall value of 70.9%. A summary of the segmentation performance comparison between the evaluated models is presented in Table 3.

Table 3. Segmentation performance comparison between the proposed YOLO-SAM workflow and Mask R-CNN evaluated on the same dataset partitions.

Model	Precision (%)	Recall (%)	mAP@0.50 (%)	mAP@ [0.50:0.95] (%)
YOLO + SAM	91.0	86.0	92.78	70.0
Mask R-CNN	88.4	70.9	95.0	64.2

The results indicate that the proposed YOLO-SAM workflow provides segmentation performance comparable to Mask R-CNN while supporting modular integration within the proposed measurement pipeline. Since sleeper spacing measurement in the proposed workflow depends directly on segmentation-derived centre locations, comparative segmentation performance provides an appropriate basis for evaluating the relative suitability of alternative models within the measurement pipeline.

5.3 Quantitative Measurement Validation Using Ground Truth Sleeper Spacing

Quantitative validation of the proposed measurement framework was performed using manually collected ground truth sleeper spacing data. A total of forty-eight consecutive sleeper spacings were measured manually under limited track access constraints to provide reference values for accuracy assessment. Ground measurements were obtained using a calibrated measuring tape with millimetre level resolution. Each value represented the centre-to-centre distance between adjacent sleepers along the track direction. Measurements were collected across both tracks to ensure representative coverage of the inspected section.

Corresponding spacing values were independently generated using the proposed UAV based pipeline. Predicted spacing values were derived from aligned sleeper centre coordinates and converted into millimetre units using the scale factor obtained from track gauge-based calibration. Predicted and reference spacing values were paired for direct comparison.

The comparison demonstrates consistent agreement between predicted and reference spacing values. Most predicted values remained within the allowable engineering tolerance range of ±10 mm, indicating reliable performance under practical inspection conditions. The observed spacing error distribution reflects the influence of uncertainties introduced across sequential processing stages illustrated in Figure 2. In particular, small positional deviations during detection influence segmentation boundary definition, which subsequently affects centre alignment accuracy, while uncertainty in rail localisation directly impacts the pixel-to-metric scaling factor. The cumulative effect of these sequential uncertainties is reflected in the final spacing error metrics summarised in Table 4.

Table 4. Summary of Quantitative Spacing Measurement Accuracy

Metric	Value
Number of samples	48
Mean Ground Truth Spacing (mm)	599.5
Mean Predicted Spacing (mm)	596.4
Mean Bias Error (mm)	−3.14
Mean Absolute Error (mm)	6.56
Root Mean Square Error (mm)	7.67
Standard Deviation of Error (mm)	6.99
Percentage Within ±10 mm (%)	93.75

5.4 Discussion

The results demonstrate that the proposed geometry aware pipeline enables reliable sleeper localisation and spacing measurement from UAV imagery. Consistent detection of sleepers and rails under realistic field conditions provided stable inputs for segmentation and geometric processing, supporting reliable centre estimation across consecutive sleepers.

A key contribution of the framework lies in treating detection outputs as intermediate geometric representations rather than final inspection results. Bounding box detections were refined into segmentation-based masks and aligned centre points, enabling spacing to be computed from structured positional data. Comparative segmentation results support the use of the YOLO-SAM workflow, which achieved comparable accuracy to Mask R-CNN while maintaining higher recall. This improved continuity across consecutive sleepers, which is necessary for reliable spacing computation.

Quantitative validation showed strong agreement between predicted and reference spacing values under near nadir imaging conditions. The mean absolute error of 6.56 mm and root mean square error of 7.67 mm indicate millimetre scale engineering accuracy suitable for spacing verification within practical tolerance limits. The high percentage of measurements within ±10 mm further confirms that cumulative uncertainties remained controlled across the processing stages.

From an operational perspective, the framework demonstrates potential to support preliminary inspection and construction verification tasks over short track sections. By reducing reliance on manual tape measurements and enabling spacing evaluation from UAV imagery, the proposed workflow provides a scalable basis for extending automated geometric inspection to more complex track environments.

5.5 Future Work

Future work will focus on extending the dataset to multiple railway corridors with varying sleeper types, ballast conditions, and track geometries, including curved track sections. Additional validation will be carried out using independently measured field data collected under different environmental conditions to evaluate the robustness of the proposed framework. Further development will investigate integration of the workflow into routine inspection processes through automated data handling and reporting tools to support practical railway maintenance applications.

6 Conclusion

This paper presented a hybrid computer vision framework for railway sleeper detection and spacing measurement from UAV imagery, integrating a YOLOv11m-based object detector with the Segment Anything Model (SAM Vit-H) for segmentation driven geometric refinement. The study evaluated whether segmentation-refined localisation and geometric alignment could support reliable sleeper spacing measurement using track gauge as the sole physical reference.

Quantitative validation demonstrated consistent agreement between predicted and reference spacing values, with millimetre-scale engineering accuracy achieved across consecutive sleeper intervals. These results confirm that segmentation-refined localisation combined with explicit geometric alignment enables reliable spacing measurement under near-nadir UAV conditions.

The present study establishes a controlled baseline for spacing measurement using straight track sections and near-nadir UAV imagery. Extension to curved track geometries and more variable imaging conditions remains an important direction for future investigation. Overall, the findings demonstrate the feasibility of structured geometric measurement from aerial imagery for scalable railway inspection applications.

7 References

[1] Indian Railways Institute of Civil Engineering (IRICEN). Role of Sleeper Spacing in Maintenance of Points. Ministry of Railways, Government of India. On-line: https://iricen.gov.in, Accessed: 10/01/2026.

[2] Indian Railways Permanent Way Manual (IRPWM). Indian Railways Permanent Way Manual Chapter II: Maintenance of Permanent Way. Ministry of Railways, Government of India. On-line: https://indianrailways.gov.in, Accessed: 10/01/2026.

[3] Z. Peng, H. Wang, Y. Jiang, and Y. Wang. Railway infrastructure inspection using unmanned aerial vehicles: A review. Automation in Construction, 125:103599, 2021.

[4] G. Morgenthal, N. Hallermann, T. P. Kersten, and H.-G. Maas. Surveying and monitoring of railway infrastructure using unmanned aerial vehicles. ISPRS Journal of Photogrammetry and Remote Sensing, 126:1–15, 2017.

[5] Q. Li, H. Zhang, Y. Ding, Y. Sun, and Y. Wang. Vision-based inspection of railway infrastructure: A survey. IEEE Transactions on Intelligent Transportation Systems, 22(10):6477–6497, 2021.

[6] T. Ye, H. Liu, Y. Wang, and Z. Wang. Railway object detection and fault diagnosis based on deep learning: A review. Sensors, 20(3):755, 2020.

[7] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter. Deep convolutional neural networks for detection of rail surface defects. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2016.

[8] C. Alippi, E. Casagrande, F. Scotti, and F. Piuri. Composite real-time image processing for rail inspection. IEEE Transactions on Instrumentation and Measurement, 49(3):559–564, 2000.

[9] A. Kumar and R. Singh. Measurement of railway sleeper spacing using Mask R-CNN. In Proceedings of the International Conference on Decision Aid Sciences and Applications (DASA), 2022.

[10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The PASCAL visual object classes challenge,” International Journal of Computer Vision, vol. 88, no. 2, pp. 303–338, 2010.

[11] T. Y. Lin, M. Maire, S. Belongie, et al., “Microsoft COCO: Common objects in context,” in Proceedings of the European Conference on Computer Vision, 2014, pp. 740–755.

[12] J. S. Choi and J. K. Lee, “Effects of data augmentation on the Nine-Axis IMU-Based orientation estimation accuracy of a recurrent neural network,” Sensors, vol. 23, no. 17, p. 7458, Aug. 2023, doi: 10.3390/s23177458.

[13] ReportBench: Evaluating deep research agents via academic survey tasks. (n.d.). https://arxiv.org/html/2508.15804v1

[14] Ultralytics. (2025, December 12). Performance Metrics Deep Dive. Ultralytics YOLO Docs.